

Contribution à l'analyse robuste non déterministe pour les systèmes de dialogue parlé

David ROUSSEL

Laboratoire CLIPS
Univ. Joseph Fourier, BP 53
38041 Grenoble Cedex
David.Roussel@imag.fr

Patrice LOPEZ

LORIA
Univ. Henri Poincaré Nancy 1, BP 239
54506 Vandoeuvre Cedex
lopez@loria.fr

Résumé

Nous présentons une technique d'analyse robuste dans le but de relayer la décision d'un système de reconnaissance de la parole. La stratégie d'analyse proposée est fondée sur une grammaire d'arbres adjoints lexicalisée compactée et sur la mise en concurrence des différentes hypothèses du système de reconnaissance de la parole. Les problèmes de robustesse sont étudiés en considérant les interférences entre erreurs de reconnaissance de la parole et phénomènes de parole spontanée dans les dialogues homme-machine.

1. Introduction

Pour l'analyse robuste de la parole, plusieurs stratégies d'analyse ont été proposées, dont le but est de "reconnaître" une analyse unique par des méthodes sélectives (Seneff, 92), stochastiques (Schwartz *et al.*, 96) ou heuristiques (Boufaden *et al.*, 98). Ces approches ne sont pas prévues pour mettre en concurrence différentes hypothèses d'un système de reconnaissance de parole mais pour délivrer un résultat unique. Ce type de déterminisme est adopté pour des raisons de performance, et du fait des capacités de traitement limitées (souvent volontaires) des modules d'interprétation.

Dans le cadre de l'analyse robuste pour les systèmes de dialogue parlé, une stratégie de type non déterministe conviendrait mieux en sortie du système de reconnaissance :

- la tâche de décodage de la séquence de parole ne disposant pas des connaissances suffisantes pour garantir le résultat, l'analyseur ne doit pas non plus délivrer une analyse unique correspondant à une hypothèse unique de reconnaissance. Il semble au contraire préférable de ne pas écarter trop tôt certaines analyses. En effet, le coût d'une erreur de compréhension est élevé : elle engage des interactions supplémentaires pour lesquelles des erreurs potentielles sont à nouveau à envisager.
- Une stratégie qui sélectionne, parmi les n meilleures hypothèses d'un système de reconnaissance, la première hypothèse qui maximise un critère de bonne formation donné, n'est pas satisfaisante. Ce critère possède un faible pouvoir discriminant sur les hypothèses de mots concurrentes. Sur les hypothèses de phrases, le succès de cette stratégie est aussi variable : il dépend de la rareté des séquences reconnues conformes au critère de bonne formation considéré.
- Si erreur de reconnaissance il y a, c'est bien parce qu'il existe une incompatibilité entre l'énoncé et les modèles acoustiques (défaut de prononciation, bruit) ou entre l'énoncé et le modèle de langage qui guide la reconnaissance de la parole (mot hors vocabulaire, construction mal représentée dans le corpus qui a servi à l'apprentissage du modèle de langage, etc.). Il n'est donc pas toujours pertinent de se fier au rang des hypothèses délivrées par le système de reconnaissance.

Du fait d'un parcours d'espaces de recherche très vastes, les stratégies déterministes sont préférées. Les méthodes sélectives sont ainsi largement expérimentées pour leur capacité à retrouver l'information que l'on suppose contenue dans une variété de messages.

L'approche expérimentée par (Lavie, 96)¹ comme celle de (Lang, 89)² se restreignent de leur côté à détecter la proximité entre une analyse robuste et différentes dérivations standard. Les heuristiques proposées par (Boufaden *et al.*, 98) donnent un moyen de trouver une solution qui s'écarte de façon minimale d'un ensemble de contraintes syntaxiques génériques, mais elles nécessitent la définition de conditions d'applications complexes. Un ordonnancement complexe est nécessaire car l'application des règles n'est pas régie par une combinatoire syntaxique, et surtout n'est pas monotone. Le traitement du bruit est par exemple destructif.

Enfin, les méthodes stochastiques constituent un apport important : leur intégration dans un algorithme d'analyse syntaxique permet d'exploiter des contraintes de la langue et du domaine, contrairement à d'autres modélisations stochastiques où toute combinaison est rendue possible. Pour réaliser une analyse stochastique robuste, la grammaire de référence est relâchée, de façon à continuer à rechercher une analyse probable³. Toutefois, si plusieurs analyses sont maintenues, ce n'est ni pour distinguer les analyses obtenues par recouvrement, ni les analyses peu fréquentes, ni même les analyses ambiguës. La probabilité d'une hypothèse est peu lisible car elle doit prendre en compte la vraisemblance des mots, la vraisemblance de la dérivation menée jusque là et une estimation du coût de la dérivation.

Notre travail est une contribution à l'analyse robuste, envisagée ici sous l'angle non déterministe. Cette approche donne un moyen de déterminer la probabilité de déviation d'une analyse. Elle est également complémentaire des travaux en reconnaissance de parole spontanée, qui intègrent la détection des déviations par des indices acoustiques, prosodiques et lexicaux au niveau du processus de reconnaissance (Heeman *et al.*, 96).

La difficulté de notre travail est d'envisager une analyse d'hypothèses de reconnaissance dans des contextes qui présentent des ellipses et des reprises — phénomènes oraux parmi les plus représentés dans notre corpus de travail⁴. Cette analyse constitue ici un moyen de conserver un pouvoir de discrimination entre les différents types de déviation. Pour éviter une application séquentielle d'heuristiques et des choix destructifs à la suite d'une première étape d'analyse, nous faisons collaborer des règles de dérivations standards et des règles spécifiques. Les premières déterminent les dérivations partielles maximales. Les secondes procèdent à des rattrapages locaux non déterministes en manipulant les mêmes représentations que les premières. Le langage de l'application est décrit par une grammaire d'arbres adjoints lexicalisée. Pour obtenir une analyse robuste efficace sur différentes hypothèses d'un système de reconnaissance de la parole, (Van Noord *et al.*, 98) proposent d'appliquer des buts sous-spécifiés et deux types de contraintes sur les bornes des îlots obtenus. Nous examinons dans la section suivante d'autres techniques, dédiées aux grammaires d'arbres lexicalisées. Nous présentons ensuite des exemples et quelques résultats.

2. Analyse par îlots fondée sur une grammaire d'arbres adjoints lexicalisée

2.1 Représentation des îlots et notion de parcours connexe

¹ L'algorithme proposé par (Lavie, 96) autorise un certain nombre de sauts dans la séquence à analyser.

² Le paradigme étudié par (Lang, 89) est celui des grammaires à règles pondérées.

³ L'application des probabilités permet de désambigüiser une analyse mais ne donne pas en soi un mode d'analyse robuste (sauf à considérer une grammaire très surgénérative contrôlée par les probabilités des règles).

⁴ L'application considérée est la recherche de programmes télévisés au moyen d'un système de dialogue parlé.

Pour réaliser une analyse robuste, l'algorithme d'analyse manipule des représentations linéaires d'arbres élémentaires sous forme d'automates.

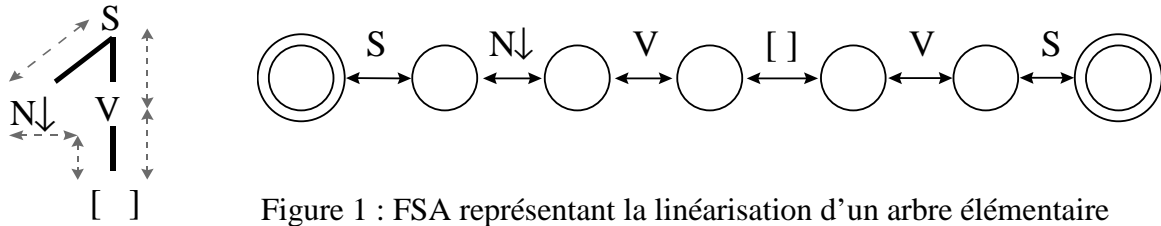


Figure 1 : FSA représentant la linéarisation d'un arbre élémentaire

La linéarisation d'un arbre élémentaire peut être représentée à l'aide d'un Automate d'Etats Finis (FSA) comme indiqué figure 1. On peut mener une analyse par *chart* où chaque *item* représente un îlot et correspond au 7-uplet suivant :

item : (*indice gauche*, *indice droit*, *état gauche*, *état droit*, *indice gauche du nœud pied*, *indice droit du nœud pied*, *état étoile*)

Les deux premiers indices sont les limites de l'îlot sur la chaîne à analyser, les deux premiers états correspondent à la position d'extension maximale respectivement gauche et droite de l'îlot. Nous représentons également, lorsque c'est nécessaire, deux indices supplémentaires pour noter la position du nœud pied d'un arbre auxiliaire englobant. L'état *étoile* correspond au nœud où l'adjonction de l'arbre auxiliaire englobant a été prédite.

En considérant des états d'analyse référant à une linéarisation d'un arbre élémentaire, nous exploitons la notion de *parcours connexe* pour définir une représentation locale associée à chaque îlot. En s'appuyant sur un FSA comme ci-dessus, on définit un *parcours connexe* comme une partie de cet automate correspondant à la liste de nœuds parcourus successivement jusqu'à rencontrer un nœud pied, de substitution ou racine (transitions incluses) ou une ancre (transition exclue). Un *parcours connexe* est un niveau de granularité intermédiaire dans la représentation d'un arbre linéarisé. Considérer des *parcours connexes* facilite la prise en compte de la topologie des arbres élémentaires et permet de se concentrer sur les nœuds significatifs pour un attachement comme expliqué dans (Lopez, 98).

Lorsqu'aucune analyse connexe ne couvre l'ensemble de l'énoncé, le résultat d'analyse correspond à des îlots et des positions d'états indiquant un arrêt sur leurs *parcours connexes*. Un point intéressant de cette représentation est que les *parcours connexes* indiquent les attentes à gauche et à droite des îlots reconnus. Cette représentation des analyses partielles facilite l'écriture de règles de rattrapage d'analyse. Nous présentons ces mécanismes sous forme de règles d'inférence. Nous notons \Rightarrow^* la fermeture transitive de la relation de dérivation entre deux items : si $i_1 \Rightarrow^* i_2$ alors l'*item* identifié par i_2 peut être obtenu de i_1 par application d'un ensemble de dérivation c'est-à-dire ici de règles d'inférence. Nous notons d'autre part avec la marque \uparrow un nœud racine. Γ_d est le *parcours connexe* passant par l'état σ_d , *suivant*(Γ) donne le premier état du *parcours connexe* après Γ selon un *parcours gauche-droite*. Enfin *tête*(Γ) (resp. *queue*(Γ)) donne la première (resp. dernière) transition (notée simplement par la catégorie correspondante) à droite (resp. gauche) de l'état le plus à gauche (resp. le plus à droite) du *parcours connexe* Γ .

Une optimisation permet de compenser le désavantage (*i.e.* la redondance) de la lexicalisation complète de la grammaire. Elle consiste à partager les sous-structures redondantes de la grammaire d'arbres, ce qui permet de parcourir une seule fois les sous-parties communes qui correspondent à des actions communes de l'analyseur. Ceci est obtenu en minimisant l'ensemble des FSA concernés en un unique FSA simulant l'ensemble de la grammaire suivant des techniques proche de (Evans & Weir, 98). La différence essentielle avec ces travaux se situe au niveau de l'automate considéré qui est ici une linéarisation des arbres, permettant d'appliquer une stratégie d'analyse quelconque sur des *parcours connexes*.

2.2 Exemple de description des reprises

Nous nous limiterons au cas des reprises au sein d'un même énoncé⁵. Lorsqu'elles sont accompagnées d'ellipses (cf. paragraphe 2.3) ou de marques venant souligner leur fonction (marques d'acquiescement, d'hésitation, etc.), les reprises considérées posent quelques problèmes aux techniques de détection par *pattern matching*, plus appropriées à la détection de répétitions ou d'ajouts (ex : « I want a film, an historical film »). Une heuristique peut en particulier induire en erreur l'interprétation, certaines précisions pouvant alors être confondues avec des corrections, comme dans (1) et (2) :

« Do you have some informations about that film a story » (1)

« I want a film yes a children's comedy » (2)

La définition de l'auto-réparation donnée dans (Cori *et al.*, 97) stipule que la partie droite de la structure interrompue (la partie droite de l'arbre dérivé au point d'interruption d'analyse) doit correspondre syntaxiquement avec la partie gauche de l'élément de reprise (structure de l'îlot adjacent). Les règles suivantes expriment cette condition dans le système déductif introduit. Pour décrire ces phénomènes, (Cori *et al.*, 97) modifient l'algorithme d'analyse pour éliminer l'élément repris. Ici, la règle (a) consiste à éliminer l'élément repris de l'analyse et la règle (b) à intégrer l'élément de reprise. L'ambiguïté des marques, comme les erreurs de reconnaissance sur ces marques, sont gérées par une règle d'absorption.

(a) Règle 1 pour l'auto-réparation :

$$\frac{(i, j, \sigma_L, \sigma_R) \quad (j, k, \sigma'_L, \sigma'_R)}{(i, k, \sigma_L, \sigma'_R)} \left(\begin{array}{l} \exists i = (v, w, \sigma''_L, \sigma''_R) \in \Delta, i \Rightarrow^* (i, j, \sigma_L, \sigma_R) \wedge \\ (\exists X \in \Gamma''_R \wedge \text{tête}(\Gamma'_L) = X *) \vee \\ (\text{queue}(\Gamma''_R) = X \downarrow \wedge \text{tête}(\Gamma'_L) = X \uparrow) \end{array} \right)$$

(b) Règle pour la précision :

$$\frac{(i, j, \sigma_L, \sigma_R) \quad (j, k, \sigma'_L, \sigma'_R)}{(i, k, \sigma_L, \sigma'_R)} \left(\begin{array}{l} \exists X, X \in \Gamma_L \wedge X \in \Gamma_R \wedge \\ X \in \Gamma'_L \wedge X \in \Gamma'_R \end{array} \right)$$

2.3 Exemple de description des ellipses

Par opposition aux omissions réalisées par un système de reconnaissance de la parole, et aux constructions grammaticales elliptiques (Sarkar & Joshi, 96), les ellipses sont identifiables en contexte. Nous avons étudié les ellipses dans les réponses à une demande d'information, dans les requêtes (explicites ou implicites), et dans les actes de confirmation.

Les situations de dialogue où l'on rencontre fréquemment des ellipses sont les réponses des utilisateurs aux demandes de confirmation ou d'informations. Cet enchaînement d'actes de dialogue crée un focus d'attention sur un référent qui favorise des formes elliptiques comme :

Système: « Would you like to watch the film or abroad ? » (3)

Utilisateur: « Abord » (4)

Le traitement des énoncés elliptiques et des énoncés comportant des omissions sont différents. Dans le premier cas, l'analyse doit identifier le référent dont est fait l'ellipse à partir du contexte (historique du dialogue, entité saillante, etc.). Dans le second cas, l'analyse doit tenter une réparation à partir des hypothèses de mot concurrentes ou de tables de

⁵ (Thévenon, 89) indique des moyens simples de bloquer les reprises possibles de l'utilisateur sur deux tours de parole. La machine ponctue par exemple certaines de ses interventions d'une question comme « ceci vous satisfait-il ? ».

confusions appliquées au voisinage du point d'arrêt de l'analyse. Or, il existe des interférences entre erreurs de reconnaissance, ellipses et même reprises : les reprises (hors corrections totales) sont propices aux ellipses du fait de la contiguïté entre l'élément de reprise et l'élément de la phrase de base. Une substitution du déterminant *a* (les substitutions ou omissions de mots monosyllabiques sont très courantes) dans la reconnaissance de l'énoncé (5) donne à cet énoncé le même statut syntaxique que l'énoncé (6) où le second constituant de type nominal est une reprise de celui qui le précède.

« What is this film ~~a~~/the science fiction (film) ? » (5)

« I want to watch this film ~~hum~~ the classical (film) ? » (6)

Les intuitions linguistiques sur les phénomènes de reprise sont ici mises en défaut : ce n'est ni le fait que le second constituant comporte une ellipse d'un constituant présent dans la structure précédente, ni le déterminant défini, ni la répétition (éventuelle) du substantif *film* qui permet d'affirmer que (5) est ou n'est pas une reprise. Le risque est d'interpréter (5) comme une demande d'information sur un film de science fiction et non une demande de confirmation sur le genre d'un film. Sans la connaissance du contexte et du résultat de l'analyse des autres hypothèses de reconnaissance, l'ambiguïté ne doit pas être écartée.

La règle (c), dédiée au traitement des ellipses, modifie les représentations manipulées par l'analyseur au même titre que l'adjonction et la substitution. Cette règle est complémentaire d'autres règles d'annihilation des contraintes de substitution (cas de l'énoncé 4). L'analyseur peut ainsi compléter les dérivations par d'autres règles de réduction.

La déclarativité des règles est importante. Les constructions avec métonymie sont encore un autre exemple de construction orale nécessitant des traitements locaux intermédiaires :

« I would like to watch the twenty o'clock on bbc one » (7)

Dans (7), le rattachement du groupe prépositionnel final n'est possible qu'après avoir reconnu *the twenty o'clock* comme un groupe nominal. Il est nécessaire de détecter que *twenty o'clock* est elliptique dans ce contexte et doit être transcategorisé en un substantif. C'est le rôle de (c). On note que cette règle reste pertinente si une auto-réparation intervient et/ou si les îlots sont séparés par une marque d'hésitation en appliquant (b) et une règle d'absorption.

(c) Règle d'ellipse pour une tête sur la droite (transcategorisation) :

$$\frac{(i, j, \sigma_L, \sigma_R) \quad (j, k, \sigma'_L, \sigma'_R)}{(i, k, \sigma_L, \sigma'_R)} \left(\begin{array}{l} queue(\Gamma_R) = X * \wedge \\ ((tête(\Gamma'_L) = X \downarrow \vee \quad tête(\Gamma'_L) = X *) \wedge queue(\Gamma'_R) = X \uparrow) \end{array} \right)$$

3. Quelques résultats

Les règles d'analyse ont été implantées en Java et mises en œuvre en complément d'une analyse LTAG ascendante (Lopez, 98). Un corpus de dialogues homme-machine en français obtenus par une expérience de Magicien d'Oz a servi de corpus test. Il s'agit d'un corpus de commandes vocales nommé GOCAD. Le corpus d'interrogation sur les programmes télévisés, plus complexe, a servi pour la mise au point des règles. Nous avons fait l'hypothèse que les phénomènes considérés ne sont pas spécifiques à ce corpus français.

Nb d'énoncés	Nb de mots	Nb moyen de mots/énoncé	% d'énoncés complètement analysés	Nb moyen d'analyses par énoncé
861	5535	6,4	78,3	2.0

Tableau 1 : Résultats globaux d'analyse du corpus Gocad

La grammaire utilisée était très restreinte (529 formes fléchies et environ 80 motifs d'arbres élémentaires). Le tableau 1 présente les résultats généraux obtenus à la suite d'une première passe. Le tableau 2 donne les résultats de rattrapage du système selon les différents phénomènes oraux. Le tableau 3 présente les résultats d'une compaction par automate d'états

finis d'une famille de constructions verbales. L'automate minimalisé permet une factorisation moyenne des états de l'ordre d'un facteur 20.

Enoncés déviants	Avec hésitation	Répétitions	Reprises	Ellipses
Occurrence	123	28	22	15
Résultat (%)	79,6	78,5	63,6	46,7 (détection)

Tableau 2 : Résultats obtenus suivant une analyse robuste

<i>Automates</i>	<i>Nb arbres</i>	<i>Nb d'états</i>	<i>Nb transitions</i>	<i>Nb arbres par état</i>
<i>divisés</i>	28	273	245	1
<i>factorisés</i>	28	13	23	21.84

Tableau 3 : Exemple de compaction d'une famille d'arbres élémentaires

4. Conclusion

Pour séparer les cas d'erreur de reconnaissance, d'incomplétude de la grammaire ou de phénomènes agrammaticaux, nous proposons de circonscrire les constructions possibles et de représenter les phénomènes oraux en dehors de la grammaire, sans recourir toutefois à des heuristiques. La séparation réside en partie dans les règles d'analyse spécifiques appliquées. Des contraintes de plus haut niveau, faisant appel à une représentation du contexte, permettront d'affiner les prédictions. Différentes régularités sur les constructions linguistiques en fonction des actes de langage sont observées par (Thévenon, 89). Ces préférences doivent être intégrées à l'analyse lexicale. L'intégration de contraintes sémantiques est aussi à explorer. Dans les grammaires d'arbres adjoints synchrones, la relation syntaxe / sémantique est directement codée dans le lexique sous forme de relations entre les nœuds syntaxiques et les éléments d'une structure sémantique. Un tel mécanisme offrirait un contrôle efficace dans un domaine d'application restreint. Il permettrait également de déterminer directement le rôle de certains îlots d'une analyse partielle.

Références

- BOUFADEN N., DESLILE S. & MOULIN B. (1998), "Analyse syntaxique robuste de dialogues retranscrits : peut-on vraiment traiter l'oral à partir de l'écrit", *TALN'98*, Paris.
- CORI M., DE FORNEL M. & MARANDIN J.-M. (1997), "Parsing Repairs", *Recent advances in natural language processing*, R. Mitkov and N. Nicolov et J. Benjamins eds.
- EVANS R. & WEIR D. (1998), "A structure-sharing parser for lexicalized grammars", *COLING-ACL'98*, Montréal.
- HEEMAN P., LOKEM-KIM K. & J. F. ALLEN (1996), "Combining the detection and correction of speech repairs", *ICSLP'96*, Philadelphie.
- LANG B. (1989), "A Generative View of Ill-formed Input Processing", *ATR Symposium on Basic Research for Telephone Interpretation (ASTI)*, Kyoto, 1989.
- LAVIE A. (1996), *GLR*: A Robust Grammar-Focused Parser for Spontaneously Spoken Language*, Ph.D. thesis, Carnegie Mellon University.
- LOPEZ P. (1998). "Analyse guidée par la connexité de TAG lexicalisées", *TALN'98*, Paris.
- SARKAR A. et JOSHI A. (1996). "Coordination in Tree Adjoining Grammars: Formalization and Implementation". *COLING'96*, Copenhagen.
- SENEFF (1992). "Robust Parsing for Spoken Language Systems", *ICASSP'92*, San Francisco, Ca.
- SCHWARTZ R., MILLER S., STALLARD D., MAAKHOUJ J. (1996). "Language Understanding Using Hidden Understanding Models", *ICSLP'96*, Philadelphie.
- THÉVENON E. (1989), "Le couple question-réponse: satellites et reprises", *Analyse linguistique d'un corpus*, pp. 135-147, Tome 2, Presse Universitaires de la Sorbonne Nouvelle.
- VAN NOORD G., GOSSE BOUMA, KOELING ROB & NEDERHOF MARK-JAN (1998). "Robust Grammatical Analysis for Spoken Dialogue Systems", *Natural Language Engineering* 1(1), pp. 1-48, Cambridge University Press.